Silverfort

AI Agent Security

AI agents are transforming enterprise operations. Silverfort ensures they don't transform your attack surface, keeping your AI initiatives protected as they scale.

Silverfort continuously discovers, monitors, and protects every AI agent interaction – enabling organizations to accelerate AI adoption securely, with confidence and control. As the industry's first unified identity security platform, Silverfort holistically protects not only AI agents, but also their human owners, and the non-human identities they rely on – all within a single control plane.

The challenge: Autonomous agents, uncontrolled access

AI agents have become trusted actors across enterprise environments, accessing data, triggering workflows, and making decisions. The problem? Most organizations have no visibility into which agents exist, who owns them, or what they can access.

- → Over-privileged access : Agents inherit human-level permissions and are privileged by design - a perfect setup for lateral movement.
- → Lack of visibility & ownership : No clear inventory of agents, their owners, access permissions and relationships.
- → Autonomy without control: Agents act independently with no guardrails on access or actions, opening paths for data exfiltration and abuse.
- → No lifecycle management : Set-and-forget agents leave behind stale credentials attackers can exploit.

"By 2028, 25% of enterprise breaches will be traced back to AI agent abuse involving both external attackers and malicious insiders." – **Gartner**

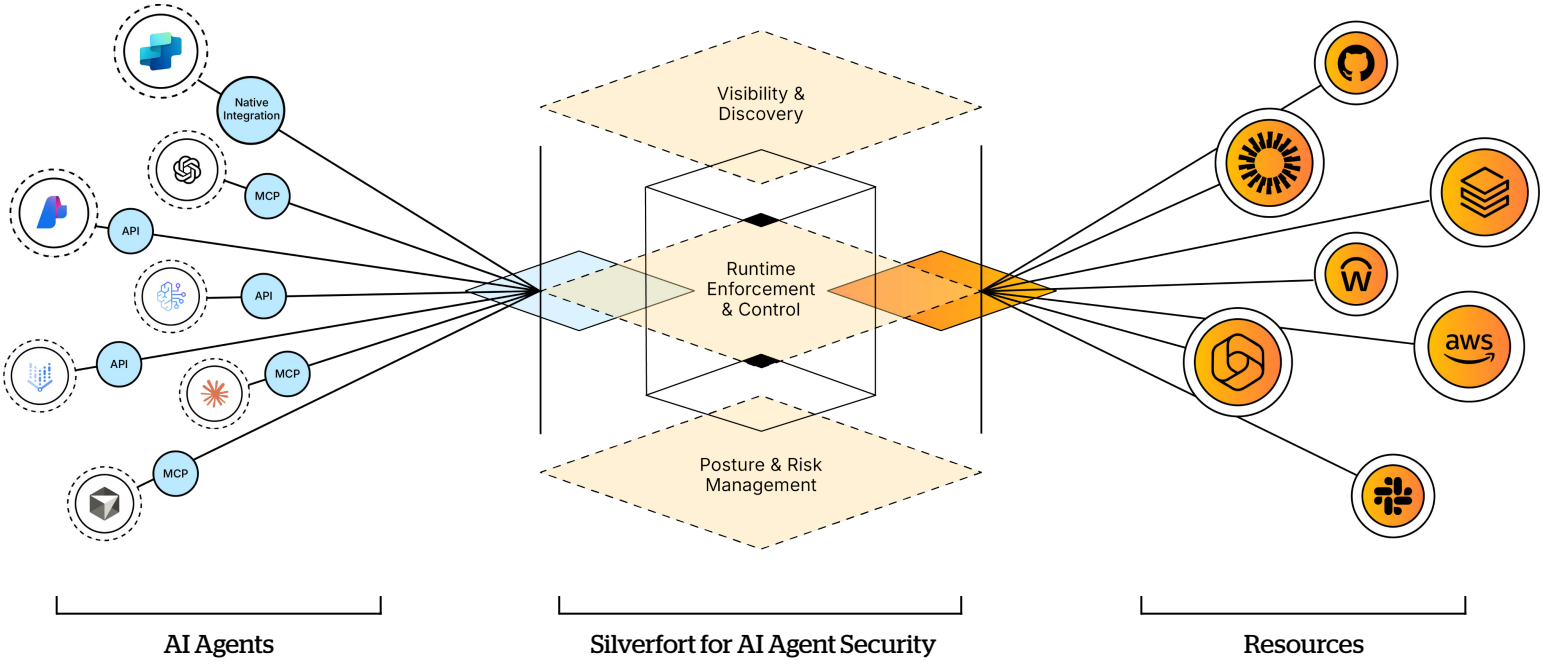
Regain control with real-time visibility and policy enforcement

Gain full visibility and context into every AI agent. Understand what each agent can access, who owns it, and how it behaves-so you can trace activity and investigate faster.

Reduce risk and enforce control by automatically detecting over-privileged or risky agents, applying least privilege and dynamic access policies, and blocking unauthorized activity.

Stop AI agent overreach by monitoring and securing every AI-driven action to prevent unauthorized access, misuse, data leakage, lateral movement, and privilege escalation.

How it works

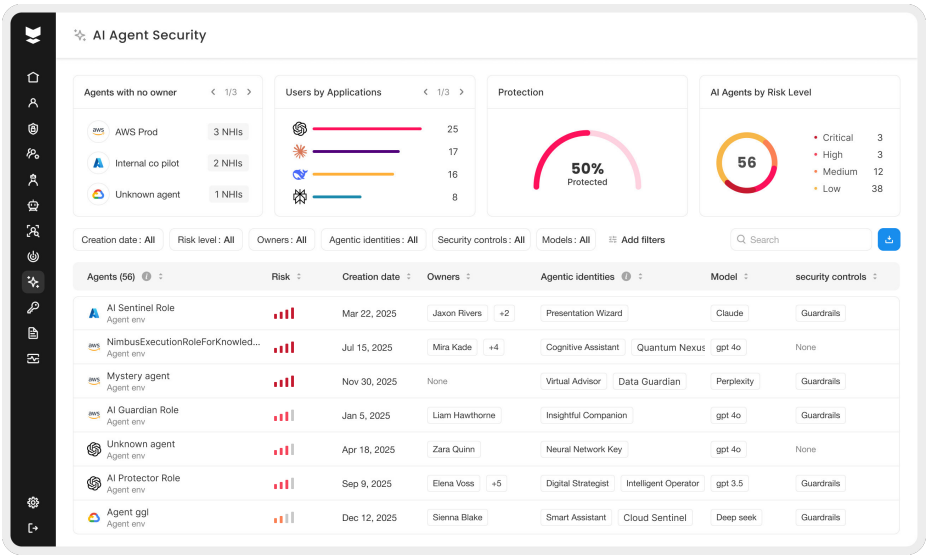


Silverfort continuously discovers, monitors, and protects AI agents across cloud, SaaS, and hybrid environments. Silverfort follows a three-step approach to unify visibility, identify risks, and enable proactive protection in real time—all without disrupting automation:

Step 1 – Discover and map: Automatically identify every AI agent across AWS, Azure, GCP, Entra ID, and SaaS environments. Map each agent to its human owner and any associated non-human identities to establish full accountability and context.

Step 2 – Assess and prioritize risks: Analyze each agent's privileges, activities, and access paths to uncover excessive permissions, supply chain exposures, and anomalous behaviors that may indicate misuse or risk.

Step 3 – Enforce policies at runtime: Apply real-time access controls through an MCP Gateway or native integrations to enforce least privilege and instantly block any unauthorized actions. Every agent operates only within its approved boundaries.



About Silverfort

Silverfort secures every dimension of identity—human, machine, and AI—across cloud and on-prem environments. We are the first to deliver an end-to-end identity security platform that is easy to deploy and doesn't disrupt business operations, resulting in better security outcomes with less work. Discover every identity across every environment, analyze exposures to reduce your attack surface, and enforce security controls at runtime to stop lateral movement, ransomware, and other identity threats.